

Assessing the Value of Inductive and Deductive Outcome Measures in Community-Based Programs: Lessons from the City Kidz Evaluation

Rich Janzen

*Centre for Community Based Research/Renison University College,
University of Waterloo
Kitchener, Ontario*

Nghia Nguyen

*Centre for Community Based Research
Kitchener, Ontario*

Alethea Stobbe

*Centre for Community Based Research
Kitchener, Ontario*

Liliana Araujo

*Centre for Community Based Research
Kitchener, Ontario*

Abstract: *Evaluators of community-based programs frequently need to decide whether to adopt an inductive or deductive approach in developing quantitative outcome measures. This article explores this issue using a case example of a child anti-poverty program called City Kidz. Its recent evaluation combined an inductive and deductive approach to develop a survey. The article describes the City Kidz evaluation and its survey before assessing the value of the survey, considering internal consistency and various aspects of validity. The article concludes with a discussion about the factors that helped and hindered the appropriateness of the survey in light of the inductive and deductive approaches used.*

Keywords: *community-based research, inductive versus deductive approaches, outcome evaluation, survey design*

Résumé : *Les évaluations de programmes communautaires nécessitent fréquemment de choisir entre une approche inductive et une approche déductive pour développer des indicateurs de résultats quantitatifs. Pour explorer cette question, cet article examine une étude de cas d'un programme sur la pauvreté enfantine nommé City Kidz. L'évaluation de ce programme a employé à la fois une approche inductive et*

Corresponding author: Rich Janzen, Renison University College, University of Waterloo, 73 King St. W., Suite 300, Kitchener, ON, Canada, N2G 1A7, (519) 741-1318, ext. 233; rich@communitybasedresearch.ca

déductive lors du développement d'une enquête. Cet article décrit l'évaluation ainsi que son enquête, puis évalue la cohérence et la validité de cette dernière. Cet article conclut en discutant des facteurs qui ont facilité ou nuit à la pertinence de l'enquête par l'entremise d'une approche inductive et déductive.

Mots clés : recherche communautaire, approches inductives versus déductives, évaluation des résultats, conception de l'enquête

INTRODUCTION

Evaluating program outcomes is a relatively common evaluation activity at not-for-profit organizations across Canada (Cousins et al., 2008). A frequent challenge that these community-based organizations face is how best to measure program outcomes. It is not unusual for those assigned to do the evaluation to wonder whether it is best to build a customized measurement tool from the “ground-up” in keeping with the uniqueness of the local context, or if it is better to draw on existing measurement tools, developed by others, that have been previously validated. In short, the challenge is whether to adopt an inductive versus deductive approach in developing outcome measures that are appropriate to the specific program being evaluated.

This decision confronting the evaluation team—developing an outcome measurement tool inductively or deductively—appears straightforward on the surface. However, there is growing recognition of the challenges inherent in measuring program outcomes. Challenges include the contextualized nature of program outcomes (Leatherdale, 2009), the complexity of client needs (Bishop & Vingilis, 2006), the complexity of triangulating information from multiple and divergent stakeholder perspectives (Obeid & Lyons, 2010), and the limited availability of outcome measurement tools, in contrast to tools measuring program improvement (Yohalem & Wilson-Ahlstrom, 2010). In addition, Carman (2007) and Sehl (2004) note that community-based programs often lack the evaluation capacity to adequately measure outcomes. Capacity challenges include an ill-defined program theory, limited evaluation knowledge, uncertainty as how to measure divergent outcomes among diverse clients, and the excessive burden of evaluation on existing staff workloads. All of these factors add a level of complexity when deciding whether to take an inductive or deductive approach to outcome measurement, and often leave community-based programs struggling with evaluation.

This article explores the inductive versus deductive distinction in quantitative outcome measurement using a case example of a child anti-poverty program called City Kidz. City Kidz is a community-based program experiencing many of the outcome measurement challenges described above. Its recent evaluation combined an inductive and deductive approach to developing quantitative measures of child outcomes, and drew on both the program's own change model and external conceptual models when developing a survey tool. Adopting both approaches

when developing a single survey tool provides a favourable opportunity to assess the value of each approach. Although the inductive/deductive decision is not unique to community-based settings, the City Kidz example does allow consideration of community-based contextual factors that led to the success and limits of each approach.

We begin the article by briefly providing background to the City Kidz evaluation and its outcome survey tool. Next, we describe the method in which the value of the various outcome measures were assessed, considering both the internal consistency of survey items and various aspects of validity. The results of the assessment are then presented before discussing factors that helped and hindered the success of inductive and deductive approaches. The article concludes by reflecting on the implications of City Kidz lessons for scholars and practitioners of outcome evaluation.

BACKGROUND: CITY KIDZ AND ITS OUTCOME SURVEY TOOL

City Kidz is a faith-based not-for-profit organization that has been working for nearly 20 years to impact the lives of children in the lowest income neighbourhoods of Hamilton, Ontario. World Vision Canadian Programs is a major funder through its Partners to End Child Poverty (PECP) program. The City Kidz theory of change is summarized in its program logic model (see [Figure 1](#)). Three levels of activities are shown at the top: (a) spiritual discipline activities that help the organization to discern God's leading, (b) activities for groups of children, centred on the Saturday theatre shows that seek to emulate Walt Disney's ability to entertain and inspire wonder, and (c) activities for individual children via weekly home visits that draw inspiration from Mother Theresa's example of humility, nurture, and care. The theory of change suggests that prolonged involvement increases the likelihood of positive outcomes for child participants.

Three types of anticipated outcomes appear in the program logic model: faith, resiliency, and child well-being. The faith outcomes acknowledge the three main messages of City Kidz programming (God created me; God loves me; God has a plan for my life). Resiliency outcomes draw on the resiliency theory developed by [Resiliency Canada \(2001\)](#), but are grounded in City Kidz's own understanding of resiliency. The resiliency outcomes are themselves clustered into three groupings: internal, relationship, and action. Internal outcomes recognize the need for children to increase their internal capacity as healthy individuals and mirror many of the internal strengths found in resiliency theory. Relationship outcomes recognize the need for children to increase supportive influences in their life and mirror selected external strengths found in resiliency theory (related to family and peers). The action outcomes portion recognizes the need for children to increase how they imagine and pursue positive activity and relies less on formal resiliency theory than on City Kidz's own theory of change. Hope outcomes are emphasized as being central to City Kidz programming and lead to increased child safety, health, and education ([Janzen, Araujo, and Stobbe, 2013](#)).

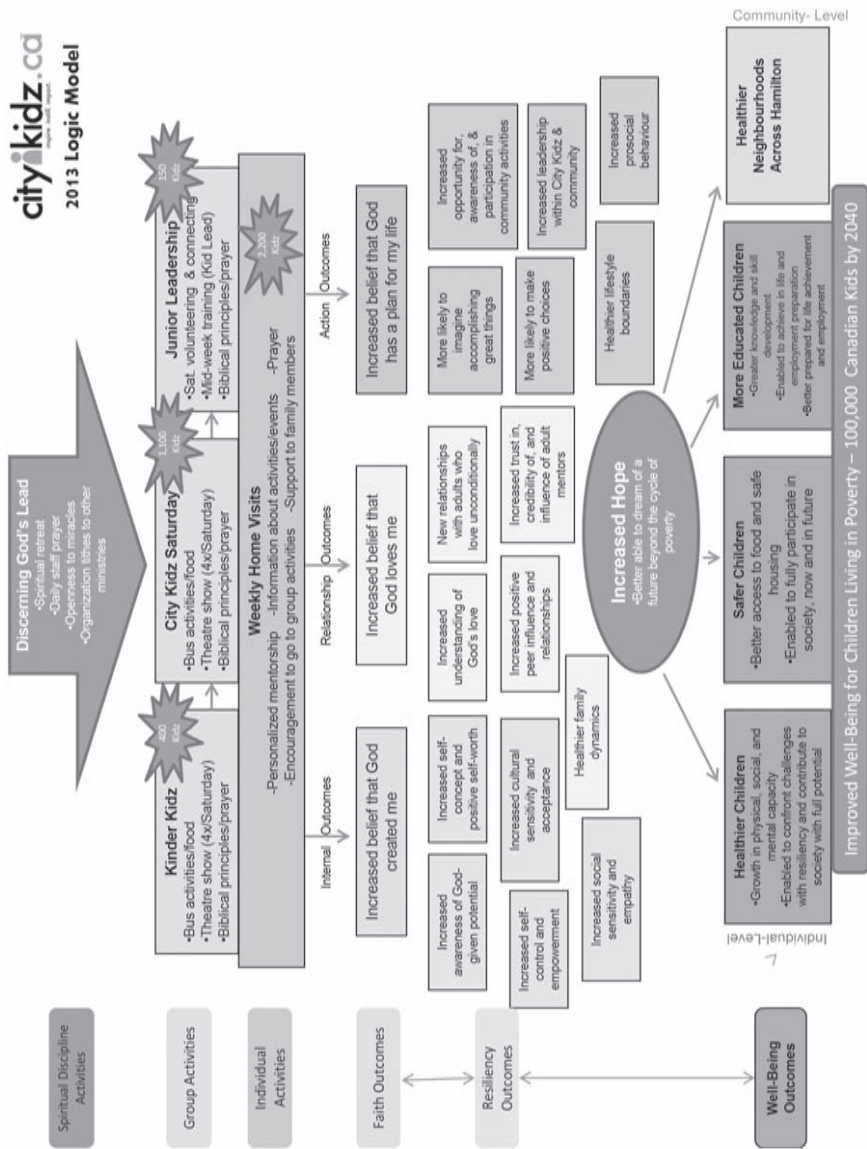


Figure 1: City Kidz Program Logic Model

The evaluation of City Kidz was completed in 2012–2013. The evaluation used a community-based approach to engage diverse stakeholders (Israel, Schulz, Parker, & Becker, 1998; Ochocka & Janzen, 2014). Concrete mechanisms implemented to support this approach included a cross-stakeholder steering group to guide each step of the process, a participatory process in developing the evaluation framework (including main research questions, program logic model, measurement matrix, toolkit, detailed work plan, and staff/volunteer training manual), the training of internal staff/volunteers as interviewers, ongoing feedback of evaluation findings, and cross-stakeholder input into recommendations (Janzen, Seskar-Hencic, Dildar, & McFadden, 2012).

Although the evaluation used multiple methods to gather information from multiple stakeholder perspectives, the focus of this article is on a survey administered to child participants. More specifically, the focus is on the 27 closed-ended questions of the survey tool dealing with child outcomes. Consistent with outcomes in the program logic model, closed-ended questions were divided into four outcome sections: (a) six internal questions beginning with “Thinking about yourself,” (b) seven relationship questions beginning with “Thinking about your relationships,” (c) seven action questions beginning with “Thinking about the things you do,” and (d) seven hope questions beginning with “Thinking about your goals.” A fifth outcome section (*faith*) was embedded within the internal, relationship, and action sections.

The bulk of the survey questions (21 of 27 items) were inductively developed, directly corresponding to the internal, relationship, action, and faith outcomes found in the program logic model. Each of these questions had five response options (*strongly agree, agree, neutral, disagree, and strongly disagree*). One inductively developed question on the topic of hope (“I have hope for my future”) had similar response options to the outcome sections above. The remaining six questions were deductively developed, based on the 6-item Children’s Hope Scale (Snyder, Cheavens & Sympton, 1997). These questions offered a choice of six responses (*none of the time, a little of the time, some of the time, a lot of the time, most of the time, and all of the time*). The individual questions (and corresponding outcome area and question design approach) are found in Table 1.

Survey participants were randomly sampled from a list of 1,347 children between the ages of 7 and 12 years who had participated in Saturday programming at least once in the last four weeks of 2012. A total of 124 children completed the survey. Parental consent was obtained before the face-to-face individual interviews with trained staff and volunteers. The sampling error was calculated at 8.3% with a 95% CI.

METHOD: HOW THE VALUES OF OUTCOME MEASURES WERE ASSESSED

In this section we clarify the inductive/deductive distinction. We begin by describing inductive reasoning and its application to outcome measurement before turning to deductive reasoning. We then outline the criteria we used to assess the

Table 1. Child Outcome Survey Questions

Outcome area	Survey items	Item design approach
Faith	1. I believe God created me	Inductive
	2. I believe God loves me	Inductive
	3. I believe God has a plan for my life	Inductive
Internal ^a	2. With God I can do great things	Inductive
	3. I think I am important	Inductive
	4. I am in control of my life	Inductive
	5. I respect other people	Inductive
	6. I accept people who are different than myself	Inductive
Relationship ^a	2. My understanding of God's love is better today	Inductive
	3. There are adults in my life who love me for who I am	Inductive
	4. I usually get along with my family	Inductive
	5. I usually get along with my friends	Inductive
	6. I have adults in my life who I trust	Inductive
	7. There are adults in my life who I look up to	Inductive
Action ^a	2. I can imagine myself doing great things	Inductive
	3. I make good choices	Inductive
	4. I go to different activities in my neighbourhood	Inductive
	5. I would like to volunteer with City Kidz someday	Inductive
	6. I try to do what is good for others	Inductive
	7. I stand up for myself	Inductive
Hope	1. I have hope for my future	Inductive
	2. I think I am doing well ^b	Deductive
	3. I am doing just as well as other kids my age ^b	Deductive
	4. I think the things I have done in the past will help me in the future ^b	Deductive
	5. I can think of many ways to get the things in life that are most important to me ^c	Deductive
	6. When I have a problem, I can come up with lots of ways to solve it ^c	Deductive
	7. Even when others want to quit, I know that I can find ways to solve the problem ^c	Deductive

^aItems begin with #2 because the first question in each of these outcome sections was a faith question. ^bAgency thinking items on the Children's Hope Scale. ^cPathways thinking items on the Children's Hope Scale.

value of inductively and deductively devised outcome measures in the City Kidz evaluation.

An inductive approach assumes a “bottom-up” line of reasoning that moves from a particular or specific premise to reach a general conclusion (Bluedorn, 1995). Applied to outcome measurement, an inductive approach assumes that sound measurement lies in the detail of the specific program. The individual items within an outcome measurement tool are developed through understanding (or “observing”) the specific outcomes the program intends to achieve, and then inferring reasonable questions based on these outcome constructs. (For example, “Child well-being as defined by this program is connected to specific outcome dimensions. Therefore a tool measuring the well-being of child participants should include questions directly related to these specific outcome dimensions.”) The value of the evaluation tool is linked to how accurately intended program outcomes are articulated and how accurately inferences were made in how questions were subsequently framed.

In contrast, a deductive approach assumes a “top-down” line of reasoning that moves from a general premise to reach a specific conclusion (Bluedorn, 1995). Applied to outcome measurement, a deductive approach assumes that sound measurement lies in a predetermined external theory. The individual items within an outcome measurement tool are developed on the basis of an empirically based theoretical framework. (For example, “Child well-being has elsewhere been found to be connected to specific outcome dimensions. Therefore a tool measuring the well-being of participants should include questions that relate to these predetermined outcome dimensions.”) The value of the evaluation tool is linked to the extent to which questions have been shown to be consistent with a generally accepted theoretical framework (i.e., previously validated) and how applicable that theoretical framework is to the program under evaluation.

Regardless of how quantitative outcome measures are developed—whether inductively or deductively—the assessment of their value is the same. This assessment is important, as the foundation of rigorous quantitative research design rests on the use of measurement tools that are metrically sound. Reliability and validity have traditionally been the key indicators in determining the soundness (or the value) of quantitative outcome measures. In fact, confirming the reliability and validity of quantitative measures is seen as foundational in assuring the integrity of study findings (DeVon et al., 2007).

Reliability refers to the consistency or stability of a measure over time (Bryman, Bell, & Teeven, 2012) and is often seen as a prerequisite to validity (Cook & Beckman, 2006). Although there are numerous ways to categorize and measure reliability, our assessment of outcome measures considered the internal consistency of survey items. Internal consistency measures whether groupings of questions within a survey tool are related to each other. In other words, high internal consistency would be demonstrated if scores measuring a single construct correlated highly (Cook & Beckman, 2006). We assessed internal consistency by calculating Cronbach’s alpha coefficient for the total score, eliminating one at a

time. A Cronbach's alpha in excess of 0.70 is usually considered to show adequate internal consistency.

Measurement validity, on the other hand, answers the question of whether the survey tool has the ability to measure the property that it intends to measure (Bryman et al., 2012). Although there are differing conceptualizations of validity, contemporary thinking suggests that all types of measurement validity fall under the broad heading of "construct validity" (Cook & Beckman, 2006; DeVon et al., 2007). This heading suggests that a tool's scores are only useful to the extent that they reflect the construct that is intended (Cook & Beckman, 2006). Under the umbrella heading of "construct" are several subcategories, each providing a building block of evidence. Measurement validity is therefore not an all-or-none proposition but triangulates support for the validity of a construct through its various subcategories (DeVon et al., 2007). The subcategories we selected, and how we measured them, are listed below.

Face validity refers to whether, at first glance, the measures appear to be valid in the opinion of people who have expertise in the area (Bryman et al., 2012). This way of assessing validity is perhaps the simplest and most intuitive, representing a good starting point for further validity exploration. Three mechanisms were implemented to assess face validity in the City Kidz evaluation. First, an evaluation steering committee assisted in developing the survey tool. This committee was made up of multiple stakeholder perspectives (including management, front-line staff, volunteers, Board members, community partners, and funder). Committee members first helped to build program theory by responding to drafts of a program logic model. They then used their respective expertise to respond to a draft of the survey tool that was based on the outcomes of the program logic model. Second, the survey tool was pilot tested by the research team with four non-City Kidz children of similar age and differing cultural backgrounds/English-language proficiency (as per program participants). Third, reflective feedback about the survey tool was provided by the City Kidz staff and volunteers both as part of their training and as part of the evaluation after administering the tool.

Discriminant validity refers to whether the measurement tool has the ability to detect true differences between groups and to detect no difference when there is not one (Howard, 2008). Within the City Kidz evaluation, discriminant validity was linked to the length of involvement in programming. The survey was meant to help test the assumption that children with longer involvement at City Kidz would demonstrate greater outcomes (i.e., have higher scores) than those with shorter involvement. According to program theory, greater program dosage would lead to greater changes in the outcomes identified in the program logic model. If analysis showed statistical differences (through *t*-tests or ANOVA tests) between children involved for a short versus long time, then discriminant validity would be established.

Convergent validity refers to whether the measurement of a concept relates to a second measure of the concept that uses a different measurement technique (Bryman et al., 2012). Exploring convergent validity is recommended for tools dealing with spiritual issues, given their highly subjective nature (Parsian &

Dunning, 2009). Assessment of convergent validity was made possible within the City Kidz evaluation because several different methods were used to explore program processes and outcomes. While children were asked closed-ended questions about program outcomes through the City Kidz survey tool, several qualitative methods were also used to explore child-level outcomes: open-ended questions on the City Kidz survey tool; individual and focus groups interviews with program participants, parents, and staff/volunteers; key informant interviews with community partners and experts; and in-depth case studies of selected City Kidz participants/graduates. The qualitative data from these methods were initially coded using content analysis by individual method before developing themes that emerged across methods.

Internal structure refers to whether scores follow a pattern as predicted by the constructs—that scores intended to measure a single construct yield homogeneous results while scales intended to measure multiple constructs yield anticipated heterogeneous results (Cook & Beckman, 2006). The particular constructs that we were assessing in the City Kidz survey tool related to the inductively developed questions within the faith, internal, relationship, action, and resiliency outcome sections. (Given the low internal consistency and the low convergent and discriminant validity of the hope items, hope was not considered in the factor analysis as these items would not form a construct.)

Both exploratory factor analysis and confirmatory factor analysis were used to assess internal structure. Exploratory factor analysis (EFA) has the capacity for identifying interpretable factors that explain the covariation of the measured variables. EFA was conducted with SPSS analysis using maximum likelihood estimation, Kaiser-Meyer-Olkin value and Barlett's test, and Eigen value and scree plot. Confirmatory factor analysis (CFA) has the advantage of testing whether theoretical relationships between items and their hypothesized factors are supported by the data. More specifically, confirmatory factor analysis requires researchers to specify the exact number of factors that exist and how these factors are related to the variables being measured. Our confirmatory factor analysis used a structural equation modelling (SEM) approach with IBM SPSS Amos 22.0 software. The criteria for evaluation of CFA included standardized regression coefficients, chi-square statistics and other methods of model fit test, Eigen values, regression weights, and standardized residual covariances.

RESULTS: DETERMINING THE VALUE OF OUTCOME MEASURES

In this section we summarize the results of assessing the value of City Kidz outcome measures according to each of the criteria discussed above. We begin with reporting on the findings of reliability before reporting on the various subcategories of validity.

Reliability (internal consistency): The results of our internal consistency analysis were mixed when considering correlations of survey questions within outcome sections. This finding suggests that questions in some outcome sections hang together better as constructs than do others. It should be noted that scores of most

survey items were generally skewed in a narrow positive range. This “ceiling effect” made it more difficult to find strong correlations.

A positive finding was that Cronbach’s alpha for faith items were acceptable (.745) despite having only three questions. The combined resiliency items (i.e., all internal, relationship, and action questions) were also adequately correlated (.790). In addition, faith and resiliency were significantly correlated (at the .01 level) as per program theory ($\alpha = .419$). However, the level of correlation indicates that they are indeed distinct constructs from each other.

Less (but still not poor) internal consistency was found for questions related to action and internal outcomes (.650 and .620, respectively). The weakest reliability was found in the constructs for hope (.579) and relationships (.399). Further analysis of interitem correlations suggested that five items could be removed (or at least reworded) to increase outcome section reliability: internal item #5 (“I respect other people”); relationship items #2 (“My understanding of God’s love is better today than it used to be”), #4 (“I usually get along with my family”), and #5 (“I usually get along with my friends”); and action item #5 (“I would like to volunteer with City Kidz someday”). However, removing or altering single items on the Children’s Hope Scale was not advisable, given that the six items of this scale had been previously validated across seven studies in the United States (see Valle, Huebner, & Suldo, 2004).

Face validity: The three mechanisms we implemented to assess face validity each confirmed that, on face value, most questions seemed to be measuring the intended constructs. Members of the stakeholder committee used their stakeholder expertise to ensure that the survey items were worded in clear, simple, yet precise language, and that they appeared to be measuring what was intended according to program theory. Pilot interviews confirmed that most questions and response options seemed to be clearly understood. Interviewer feedback also confirmed that survey items appeared to be measuring what was intended and that by far the majority of participants seemed to comprehend the majority of questions. However, feedback from both pilot interviewer and interviewer flagged the six questions from the Children’s Hope Scale as somewhat more difficult to comprehend by participants.

Discriminant validity: Our analysis demonstrated that those with longer involvement in the program had statistically higher scores on inductively developed outcome questions than did those with short involvement (particularly for faith, internal, and relationship outcomes, and to a somewhat lesser extent for action outcomes). For example, participants who were active in the program for more than two years had significantly higher total scores for faith, internal, relationship, and action outcomes than those active for less than 2 years (p -value for t -test of faith items combined = .003; for internal = .002; for relationship = .004; for action = .023). In other words, length of involvement was a good predictor of many of the anticipated outcome groups found in the program logic model. Only deductively developed questions related to hope outcomes did not show many significant differences (e.g., comparing those active more than 2 years with those active less than 2 years, a p -value of 0.240 for t -test of hope items combined was

found). Further details of the statistical analyses can be found in the final evaluation report (Janzen, Araujo, Stobbe, and Nguyen, 2013).

Convergent validity: Generally, the evaluation found a great deal of consistency between findings from qualitative methods and those of the City Kidz survey tool. Most notable was how the quantitative findings from the inductively developed survey questions generally converged with the findings of the qualitative methods, where outcome questions were asked to diverse stakeholder perspectives and in a more open-ended way. Table 2 provides evidence that convergent

Table 2. Comparing Survey Items and Qualitative Themes

Outcome area	Survey items	Qualitative themes
Faith	<ol style="list-style-type: none"> 1. I believe God created me 2. I believe God loves me 3. I believe God has a plan for my life 	<ul style="list-style-type: none"> • Increased belief in and understanding of God • Stronger relationship with God
Internal*	<ol style="list-style-type: none"> 2. With God I can do great things 3. I think I am important 4. I am in control of my life 5. I respect other people 6. I accept people who are different than myself 	<ul style="list-style-type: none"> • Increased self-confidence • Increased self-worth • Increased sense of purpose in life • Better treatment of others
Relationship*	<ol style="list-style-type: none"> 2. My understanding of God's love is better today 3. There are adults in my life who love me for who I am 4. I usually get along with my family 5. I usually get along with my friends 6. I have adults in my life who I trust 7. There are adults in my life who I look up to 	<ul style="list-style-type: none"> • Increased caring and respect toward friends • Increased adults and mentors who they can trust • Increased role models • Able to confide and trust in City Kidz volunteers
Action*	<ol style="list-style-type: none"> 2. I can imagine myself doing great things 3. I make good choices 4. I go to different activities in my neighbourhood 5. I would like to volunteer with City Kidz someday 6. I try to do what is good for others 7. I stand up for myself 	<ul style="list-style-type: none"> • Better moral decision-making • Better life choices • Increased involvement in community and church settings • Increased willingness to spread goodness and positive messages to others • Increased courage to stand up for self

*Items begin with #2 because the first question in each of these outcome sections was a faith question.

validity was strong between qualitative themes and corresponding quantitative survey items. More details of how qualitative and quantitative findings converged can be found in the final evaluation report (Janzen, Araujo, Stobbe, and Nguyen, 2013).

The hope outcome section was again more problematic. Although quantitative analysis of the City Kidz survey tool found little evidence of program impact related to hope, qualitative findings found considerable evidence. However, qualitative methods suggested that involvement in City Kidz resulted in a construct of hope that was more holistic than the purely cognitive, goal-oriented survey questions based on the agency and pathway thinking dimensions of hope found in the Children's Hope Scale. In particular, qualitative constructs of hope emphasized dreaming (the ability to imagine different future options), resolve (confidence and strength to follow aspirations), and perseverance (never giving up). These somewhat alternative constructs of hope put into question the validity of the deductively developed hope questions for City Kidz.

Internal structure: The results of our exploratory factor analysis are summarized in Table 3. The analysis reveals that (a) the faith construct includes the three faith items as presented in the survey (however, these three items should no longer appear within the internal, relationship, and action outcome sections), (b) the internal construct includes all five items in the internal outcome section, (c) the relationship construct includes four items (#s 3, 4, 6, and 7) or better with three items (#s 3, 6, and 7), and (d) the action construct includes five items (#s 2, 3, 4, 6, and 7).

Confirmatory factor analysis mirrored the findings of the exploratory factor analysis. Table 4 summarizes the results of the confirmatory factor analysis. The analysis reveals that the theoretical model for the City Kidz survey tool could be considered sound and appropriate, especially when a selected few items are removed (the same items identified in the exploratory factor analysis). Consistent with City Kidz program theory, the model consists of four constructs: faith, internal, relationship, and action, with the latter three constructs forming the construct of resiliency.

Summary of Results

Results of the internal consistency and validity of survey items were mixed. On the one hand, the internal consistency and validity of the faith, internal, action, and—to a lesser extent—relationship items indicated that these items were generally sound in measuring City Kidz outcomes. In addition, faith items were found to be distinctive from (yet related to) the internal, relationship, and action constructs, while items in these latter constructs were found to make up the construct of resiliency (as per program theory). However, five individual items were seen to be problematic (most notably within the relationship section). Removing or improving these items would increase the measurement strength of the survey.

In response to these findings, the evaluation team was confident in maintaining the basic survey structure, with relatively minor changes needed in these inductively developed sections to improve the next version of the survey.

Table 3. Exploratory Factor Analysis Results

Construct of interest	KMO and Bartlett's test	Eigenvalues and scree plot	Factor loadings	Goodness-of-fit test	Reproduced correlations	Comments and conclusions
Faith construct	Significant	Suggested one-factor construct	Suggested retaining all items	N/A	Parameter estimates reproduce the sample data well	Retain all three items in one-factor construct. All items seem to capture a latent factor "Faith."
Internal construct	Significant	Suggested one-factor construct	Item 5 may go to another factor	Not significant. Data fits the model well	Parameter estimates reproduce the sample data well	Retain all five items in one-factor construct. All items seem to capture a latent factor "Internal."
Relationship construct (excl. item 2 and 5)	Significant	Suggested one-factor construct	Item 4 may go to another factor	Not significant. Data still fits the model.	Some reproduced correlations exceed a value of .05, suggesting item 4 may not a good fit.	The construct with all 7 items suggests a three-factor solution. The construct with 5 items (excluding item 2 and 5) suggests a two-factor solution. One-factor construct could have 4 items (3, 4, 6, and 7), but would be best with three items (3, 6, and 7). All three items seem to capture a latent factor "Relationship."
Action construct (excl. item 5)	Significant	Suggested one-factor construct	Suggested retaining all five items	Not significant. Data still fits the model.	Some reproduced correlations exceed a value of .05, suggesting item 3 may not a good fit.	The construct with 6 items (excluding item 5) still suggests a two-factor solution. The construct with 5 items (excluding first item and item 5) suggests a one-factor solution and an adequate representation of the data. All five items seem to capture a latent factor "Action."

Table 4. Confirmatory Factor Analysis Results

Construct of interest	Standardized regression coefficients	Chi-square and other goodness-of-fit tests	Eigenvalues	Regression weights	Standardized residual covariances	Comments and conclusions
Confirmatory Factor Analysis for separate constructs						
Faith construct (incl. three items)	.59 to .88	NA	Suggested one-factor construct	Factor coefficients are significant	All are small	Retain this construct with all three items.
Internal construct (incl. item 2,3,4,5,6)	.32 to .56 (not high loadings)	Good fit	Suggested one-factor construct	Factor coefficients are significant	All are small	Low coefficient for item 5. Consider removing item 5 from the construct.
Internal construct (incl. item 2,3,4,6)	.46 to .58 (acceptable loadings)	Good fit	Suggested one-factor construct	Factor coefficients are significant	All are small	Good for internal construct.
Relationship construct (incl. item 3,4,6,7)	.22 to .72 (low loading with item 4)	Good fit	Suggested one-factor construct	Factor coefficients are significant	Big residuals for item 4	Consider removing item 4 from the construct.
Relationship construct (incl. item 3,6,7)	.58 to .69 (high loadings)	Good fit	Suggested one-factor construct	Factor coefficients are significant	All are small	Good for relationship construct, but not good for the theory.
Action construct (incl. item 2,3,4,6,7)	.44 to .73 (acceptable loadings)	Good fit	Suggested one-factor construct	Factor coefficients are significant	Some large residuals for item 3	Good for action construct.

Confirmatory Factor Analysis for overarching theory

After testing all possible combinations of the above constructs, we found that

- Internal, relationship, and action constructs could go well together, forming a concept of resiliency.
- Faith is a good construct by itself, but would not go well with other constructs, especially action (low correlation).

The best theoretical construct for resiliency

Resiliency construct (with item 2,3,4,6 from Internal, item 3,6,7 from Relationship, item 2,3,4,6,7 from Action)	High loadings	Acceptable fit	N/A	Factor coefficients are significant	Still have some large residuals, but acceptable for a large model	Good for resiliency construct.
--	---------------	----------------	-----	-------------------------------------	---	--------------------------------

Subsequent improvements included placing the three faith items in their own separate faith section with the heading “Thinking about God.” In addition, of the five problematic questions, three would be reworded to be clearer and better matched with program theory (these questions each being connected to a stated outcome within program theory, suggesting their importance). Finally, the remaining two items would be deleted from the next version of the survey, one because of its similarity to another survey item, and one because it was seen to be speculative for survey-aged children and could better be measured by directly asking older youth.

On the other hand, the items related to the hope construct were generally found to be less valuable in measuring City Kidz participant outcomes. The bulk of these questions (six of seven) were drawn from an existing measurement tool (i.e., the Children’s Hope Scale). Results of the internal consistency and validity of these hope items were poor. Consequently, these six questions would be omitted from the next version of the survey. Instead, additional questions would be added based on qualitative themes that emerged through other evaluation methods. Specifically, qualitative hope themes (and corresponding questions) included dreaming (“I have a bright future before me”), resolve (“I know that I will be able to do what I want to do” and “I am strong enough to do what I want”), and perseverance (“I never give up when doing something important”). The additional single hope question (“I have hope for my future”) appeared to have adequate face validity and would remain in the revised hope section.

In conclusion, survey sections that were inductively developed (i.e., faith, internal, action, and—to a lesser extent—relationship) were found to be generally of value in measuring City Kidz participant outcomes. These sections used local program theory as the basis for question formulation. Contrast these sections with the hope section that was primarily developed deductively through the adoption of an existing external measurement tool. This hope section was found to be poor in reliability and validity in measuring participant outcomes.

DISCUSSION: WHAT WE LEARNED ABOUT DEVELOPING OUTCOME MEASURES

As with all community-based research, the challenge for not-for-profit organizations who wish to pursue program evaluations is to adhere to the dual criteria of conducting research with excellence while ensuring practical relevance (Ochocka & Janzen, 2014; Ochocka, Moorlag, and Janzen, 2010). For outcome measurement, *excellence* means that validity and reliability (for quantitative measures) and trustworthiness (for naturalistic qualitative measures) can be established (Bryman et al., 2012; Lincoln & Guba, 1985). Relevance in outcome measurement means that information-gathering tools are “location-based” to the extent that they produce research findings that are useful in stimulating reflective practice within the program under evaluation (Janzen et al., 2012). The City

Kidz case example provides community-based insight into how inductive versus deductive approaches to quantitative outcome measurement can help facilitate this dual goal of excellence and relevance.

In the end, our assessment found that inductively developed measures generally proved more valuable than deductively developed measures. Our take-away conclusion is not that inductive approaches to outcome measurement are therefore always preferred over deductive approaches within community-based evaluations. But for some reason the manner in which the inductive approach was implemented in the City Kidz evaluation was more fruitful than the deductive approach. The question is why. Why was the inductive approach more successful, and what were the ingredients that made it so? Conversely, why was the deductive approach not successful in this case? Below we briefly attempt to answer these questions.

Factors Facilitating the Success of the Inductive Approach

We previously stated that within inductive approaches the appropriateness of an outcome measurement tool is linked to accuracy in articulating program outcomes, and to the strength of inference in how questions are consequently framed. Regarding the articulation of program outcomes, we see two main factors leading to success within the City Kidz evaluation:

The use of a participatory process in program theory development: Although many community-based evaluations use a program theory approach to evaluation (Chen, 2005), not all do so in a participatory fashion. At the start of the evaluation, City Kidz did not have a clearly articulated theory of change. The evaluation team therefore reviewed program documents (website, funding proposals, reports, promotional material, etc.), held site visits and discussions with key program staff, and involved the cross-stakeholder steering committee in developing a program logic model. Accuracy was ensured by triangulating these various sources and facilitating mutual agreement across stakeholder perspectives (Janzen et al., 2012; Rey, Brousselle & Dedobbeleer, 2011). The outcomes in the program logic model were the basis for developing survey questions (all anticipated outcomes having at least one corresponding survey question reworded in a child-friendly way). This participatory process underscored the importance of establishing program theory validity as a precursor to measurement validity when using an inductive approach.

The inclusion of (external) resiliency theory to sharpen (local) program theory: Many of the resiliency outcomes found in the City Kidz program theory drew on resiliency theory (Resiliency Canada, 2001). At first blush, it may seem that reaching out to external theory is more conducive to a deductive than to an inductive approach. It should be noted, however, that the primary emphasis was still on building local program theory. It was not a matter adopting all aspects of resiliency theory to explain City Kidz, given that much of resiliency theory was not seen to apply to the City Kidz context. However, those parts of resiliency theory that were seen to be relevant were incorporated (i.e., many of the internal

strengths components, and some of the family and peer strengths components). External resiliency theory therefore played a secondary, supportive role in program theory development. Its benefit was in aiding concept clarification (Taylor & Lord, 1996). That is, adding resiliency theory concepts into the participatory program theory development process allowed stakeholders to develop a common (and evidence-based) language with which to articulate what they intuitively sensed to be true.

Regarding the strength of inferences in framing survey questions, we again see two main factors leading to success within the City Kidz evaluation:

The facilitation of cross-stakeholder agreement on survey design: Once again a participatory process was critical in ensuring that specific survey questions were appropriate to the City Kidz context. The steering committee first agreed on an evaluation purpose statement and main research questions, before discussing how to translate intended outcomes into specific survey questions. Committee members were encouraged to offer their critical comments and to come to agreement on how to express questions in a child-friendly way. As members of a faith-based organization, stakeholders were also encouraged to factor faith into program theory and survey question design (Janzen and Wiebe, 2010). This facilitative process accentuates survey design as a relational, not only a technical, exercise (Janzen et al., 2012).

The pilot testing of questions with children of similar characteristics: Our pilot tests confirmed that inductive inferences being made when developing survey questions seemed to be communicated in a way that made sense for children. (In contrast, the six deductively developed questions from the Children's Hope Scale were somewhat more difficult to comprehend.) Although the number of pilot tests (4) was relatively small, children were selected in an attempt to reflect some of the diversity of City Kidz participants in terms of age, cultural background, and immigrant status (i.e., ranging from recent immigrants to Canadian-born).

Factors Hindering the Success of the Deductive Approach

We previously stated that, within deductive approaches, the appropriateness of an outcome measurement tool is linked to the extent to which questions have been previously shown to be consistent with an empirically developed theoretical framework (i.e., validated), and how applicable that theoretical framework is to the program under evaluation. The bulk of the hope questions asked on the City Kidz survey were in fact consistent with the previously validated Children's Hope Scale (see Snyder, 1995; Snyder et al., 1997; Valle et al., 2004). Instead, we believe the problem lay in the applicability of Snyder et al.'s (1997) theory of hope to the City Kidz context, specifically its demographic range and its theoretically narrow understanding of hope.

A theory of hope based on a narrow demographic of children: Snyder et al.'s (1997) theory of hope was empirically developed and validated through research on children (ranging in age from 7 to 17) in various cities across the United States

(Snyder, 1995; Snyder et al., 1997). The samples primarily involved children with a variety of medical conditions, and were relatively homogeneous in terms of racial, ethnic, and socioeconomic backgrounds. Attempts to broaden the theory's generalizability involved studies within United States high schools, including schools with lower-income, African-American children (Valle et al., 2004). Still, sample demographics did not match key aspects of City Kidz participants, namely recent immigrant children and children living in poverty within a Canadian urban environment. No discussions took place during the survey design about the impact of these sampling issues on the applicability of Snyder et al.'s theoretical framework to City Kidz.

A theoretically narrow understanding of hope: At face value, the Children's Hope Scale seemed to be a reasonable starting point from which to explore the construct of hope at City Kidz. Snyder et al.'s (1997) framework provided a more nuanced understanding of hope than previously articulated by City Kidz. As well, few if any other scales were available. In the end, the scale did not prove to be helpful in shedding more light on hope at City Kidz. Rather, it was the qualitative themes (i.e., dreaming, resolve, and perseverance) generated across stakeholder perspectives that demonstrated Snyder et al.'s cognitive-oriented framework as being too narrow in describing the hope that City Kidz was striving for. This challenge was similar to that experienced by Sehl (2004), whose deductive approach to outcome evaluation limited stakeholder involvement in survey design. This resulted in an evaluation not flexible enough to fit the unique contexts of individual programs.

CONCLUSION: IMPLICATIONS FOR SCHOLARS AND PRACTITIONERS OF OUTCOME EVALUATION

This article explored the value of inductive versus deductive approaches in quantitative outcome measurement using City Kidz as a case example. Evaluating outcomes has traditionally had a deductive bias, favouring standardized outcomes tools that were previously validated (Hinkin, 1998; Kumpfer et al., 1993). Such deductively developed tools had the advantage of being quicker to develop and cheaper to implement (Desimone & Le Floch, 2004; Myers, 1999), a bonus for community-based programs with limited evaluation budgets. However, challenges to implementing predetermined tools have been noted over the years, particularly for community-based programs who lack technical know-how (Carman, 2007) and who need to know how to adapt tools to their specific settings (Goodman, 1998; Padilla & Medina, 1996). In recent years it has become increasingly common to have complementary approaches, with inductive and deductive measures triangulated in a single evaluation (Williams, 2006; Smith et al., 2014).

The City Kidz evaluation was consistent with this trend. We see our contribution to the evaluation outcome measurement literature as stressing the importance of process when developing outcome measures. The overarching

lesson we learned was that regardless of approach taken (deductive or inductive), efforts should be made to ensure the applicability of outcome measures to the unique program context, and to do so by facilitating a participatory process involving program stakeholders. Such a collaborative process increases the likelihood that the evaluation is done with research excellence (in our case determined via measurement reliability and validity), while also being practically relevant to the program itself. In fact, this dual goal of excellence and relevance could be seen to be mutual reinforcing and not a zero sum game. That is to say, pursuing reliability and validity for outcome measurement does not need to take away from, but can rather strengthen, the pursuit of practical utility, and vice versa.

Our assessment was not without its limitations. [Cook and Beckman \(2006\)](#) note two important threats to construct validity. The first threat relates to inadequate sampling of the content domain, recognizing that establishing reliability and validity is a matter of degree and requires a broad spectrum of evidence. Our assessment could have included additional subcategories of reliability and validity to bolster our conclusions. For example, test-retest reliability could have been determined by administering the survey tool to the same participant at different times. Content validity could also have been determined by having steering committee members quantitatively rate whether each survey item was an appropriate indicator of its respective construct during the drafting of the survey tool ([DeVon et al., 2007](#)). In addition, this assessment could have been completed with a larger sample of program participants to reduce random sampling error. As it stands, the existing assessment of reliability and validity can be seen as a good starting point. Evidence could be strengthened by further assessing the revised survey tool as the program continues (and potentially expands into other Canadian cities).

A second threat relates to factors which exert nonrandom influences, including bias, on scores. The potential for a coercive effect while administering the City Kidz survey could have existed, given the young age of participants and the faith-based elements of some questions. To mitigate this effect, interviewer training included how to minimize socially desirable responses. Presumably any coercive effect would have been similar for inductively developed and deductively developed survey items, the comparison of which was at the heart of this article's assessment. Still, a coercive effect cannot be entirely ruled out and could be further mitigated in future studies.

Despite these limitations, the main lessons discussed above can be seen to be transferrable to other community-based settings. The City Kidz evaluation serves as an illustration of how collaborative approaches that draw on local expertise can develop "home-grown" evaluation tools of value, whether through an inductive, deductive, or combined approach. In this way, the lessons also provide an alternative perspective to the notion put forward by evaluators such as [Carman \(2007\)](#) and [Sehl \(2004\)](#), that community-based programs lack capacity to adequately measure outcomes.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the City Kidz evaluation steering committee and City Kidz staff for their input and dedication in implementing the City Kidz evaluation, and to World Vision Canada as an active funding partner in both the City Kidz evaluation and the assessment of the survey tool.

REFERENCES

- Bishop, J. E. H., & Vingilis, E. (2006). Development of a framework for comprehensive evaluation of client outcomes in community mental health services. *Canadian Journal of Program Evaluation*, 21(2), 133–180.
- Bluedorn, H. (1995). *Two methods of reasoning: An introduction to inductive and deductive logic*. Retrieved from: http://www.triviumpursuit.com/articles/two_methods_of_reasoning.php
- Bryman, A., Bell, E., & Teeven, J. (2012). *Social research methods* (3rd Canadian ed.). Toronto, ON: Oxford University Press.
- Carman, J. G. (2007). Evaluation practice among community-based organizations: Research into the reality. *American Journal of Evaluation*, 28(1), 60–75. <http://dx.doi.org/10.1177/1098214006296245>
- Chen, H.-T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Newbury Park, CA: Sage.
- Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability of psychometric instruments. *Theory and Application*, 119, 116.e7–166.e16.
- Cousins, J. B., Elliott, C., Amo, C., Bourgeois, I., Choinard, J., Goh, S. C., & Lahey, R. (2008). Organizational capacity to do and use evaluation: Results of a pan-Canadian survey of evaluators. *Canadian Journal of Program Evaluation*, 23(3), 1–36.
- Desimone, L. M., & Le Floch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy*, 26(1), 1–22. <http://dx.doi.org/10.3102/01623737026001001>
- DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., ... Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship*, 39(2), 155–164. <http://dx.doi.org/10.1111/j.1547-5069.2007.00161.x>
- Goodman, R. M. (1998). Principles and tools for evaluating community-based health promotion. *Journal of Public Health Management and Practice*, 4, 37–47. <http://dx.doi.org/10.1097/00124784-199803000-00006>
- Hinkin T. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1(1), 104–121. <http://dx.doi.org/10.1177/109442819800100106>
- Howard, K. (2008). *Validating questionnaires*. Kestrel Consultants. Retrieved from http://kestrelconsultants.com/reference_files/Validating_Questionnaires.pdf
- Israel, B. A., Schulz, A. J., Parker, E. A., & Becker, A. B. (1998). Review of community-based research. *Annual Review of Public Health*, 19, 173–202. <http://dx.doi.org/10.1146/annurev.publhealth.19.1.173>

- Janzen, R., Araujo, L., & Stobbe, A. (2013). *An evaluation framework for City Kidz core programs*. Kitchener, ON: Centre for Community Based Research.
- Janzen, R., Araujo, L., Stobbe, A., & Nguyen, N. (2013). *Evaluation of City Kidz core programs: Final report: Final report*. Kitchener, ON: Centre for Community Based Research.
- Janzen, R., Seskar-Hencic, D., Dildar, Y., & McFadden, P. (2012). Using evaluation to shape and direct comprehensive community initiatives: Evaluation, reflective practice, and interventions dealing with complexity. *Canadian Journal of Program Evaluation*, 25(2), 61–88.
- Janzen, R., & Wiebe, D. (2010). Putting God in the logic model: Developing a national framework for the evaluation of faith-based organizations. *Canadian Journal of Program Evaluation*, 25(1), 1–26.
- Kumpfer, K. L., Shur, G. H., Ross, J. G., Bunnell, K. K., Librett, J. J., & Millward, A. R. (1993). *Measurements in prevention: A manual on selecting and using instruments to evaluate prevention programs* (U.S. Department of Health and Human Services. Centre for Substance Abuse Prevention, Technical Report 8). Rockville, MD: Government Printing Office.
- Leatherdale, S. T. (2009). Evaluating school-based tobacco control programs and policies: An opportunity gained and many opportunities lost. *Canadian Journal of Program Evaluation*, 24(3), 89–106.
- Lincoln, Y. S., & Guba, I. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Myers, A. (1999). *Program evaluation for exercise leaders*. Windsor, ON: Human Kinetics.
- Obeid, N., & Lyons, J. S. (2010). Pre-measurement triangulation: Considerations for program evaluation in human service enterprises. *Canadian Journal of Program Evaluation*, 25(2), 59–82.
- Ochocka, J., & Janzen, R. (2014). Breathing life into theory: Illustrations of community-based research hallmarks, functions, and phases. *Gateways: International Journal of Community Research and Engagement*, 7(1), 18–33. <http://dx.doi.org/10.5130/ijcre.v7i1.3486>
- Ochocka, J., Moorlag, E., & Janzen, R. (2010). A framework for entry: PAR values and engagement strategies in community research. *Gateways: International Journal of Community Research & Engagement*, 3, 1–19. <http://dx.doi.org/10.5130/ijcre.v3i0.1328>
- Padilla, A. M., & Medina, A. (1996). Cross-cultural sensitivity in assessment using tests in culturally appropriate ways. In L. A. Suzuki & J. G. Ponterotto (Eds.), *Handbook of multicultural assessment: Clinical, psychological, and educational applications* (pp. 3–28). San Francisco, CA: Jossey-Bass.
- Parsian, N., & Dunning, T. (2009). Developing and validating a questionnaire to measure spirituality: A psychometric process. *Global Journal of Health Science*, 1(1), 2–11. <http://dx.doi.org/10.5539/gjhs.v1n1p2>
- Resiliency Canada. (2001). *Youth resiliency framework: Promoting developmental strengths through resiliency assessment and development* [Pamphlet]. Calgary, AB: Resiliency Initiatives.

- Rey, L., Brousselle, A., & Dedobbeleer, N. (2011). Logic analysis: Testing program theory to better evaluate complex interventions. *Canadian Journal of Program Evaluation*, 65(3), 61–89.
- Sehl, M. (2004). Stakeholder involvement in a government-funded outcome evaluation: Lessons from the front line. *Canadian Journal of Program Evaluation*, 19(2), 37–56.
- Smith, F., Carlsson, E., Kokkinakis, D., Forsberg, M., Kodeda, K., Sawatzky, R., ... Öhlén, J. (2014). Readability, suitability and comprehensibility in patient education materials for Swedish patients with colorectal cancer undergoing elective surgery: A mixed method design. *Patient Education and Counseling*, 94(2), 202–209. <http://dx.doi.org/10.1016/j.pec.2013.10.009>
- Snyder, C. R. (1995). Conceptualizing, measuring, and nurturing hope. *Journal of Counseling and Development*, 73(3), 355–360. <http://dx.doi.org/10.1002/j.1556-6676.1995.tb01764.x>
- Snyder, C. R., Cheavens, J., & Simpson, S. C. (1997). Hope: An individual motive for social commerce. *Group Dynamics*, 1(2), 107–118. <http://dx.doi.org/10.1037/1089-2699.1.2.107>
- Taylor, A., & Lord, J. (1996). Concept clarification as a community research tool. In C. Tolman, R. van Hezewijk, F. Cherry, & I. Lubek (Eds.), *Problems of theoretical psychology*. Toronto, ON: Captus Press.
- Valle, M. F., Huebner, E. S., & Suldo, S. M. (2004). Further evaluation of the Children's Hope Scale. *Journal of Psychoeducational Assessment*, 22(4), 320–337. <http://dx.doi.org/10.1177/073428290402200403>
- Williams, K. (2006). Action inquiry into the use of standardized evaluation tools for music therapy. *Voices: A World Forum for Music Therapy*, 6(2). Retrieved from <https://voices.no/index.php/voices/article/view/252/196>
- Yohalem, N., & Wilson-Ahlstrom, A. (2010). Inside the black box: Assessing and improving quality in youth programs. *American Journal of Community Psychology*, 45(3–4), 350–357. <http://dx.doi.org/10.1007/s10464-010-9311-3>

AUTHOR INFORMATION

Rich Janzen, PhD, is Research Director at the Centre for Community Based Research (www.communitybasedresearch.ca) and Adjunct Assistant Professor at Renison University College, University of Waterloo.

Nghia Nguyen, MD, PhD, is a former Centre Researcher at the Centre for Community Based Research and is currently Data Analyst at the University of Waterloo.

Alethea Stobbe, MSW, is a Centre Researcher at the Centre for Community Based Research.

Liliana Araujo, MA, is a former Centre Researcher at the Centre for Community Based Research and is currently the Executive Assistant to the President at the Centre for International Governance Innovation.